A solved problem on RNN

November 5, 2024

By Suneesh Jacob

Problem

An RNN model is trained with some dataset and trained weights and biases and the architecture of the model are given below.

The vocab list is

$$\text{vocab} = \begin{bmatrix} \text{am} \\ \text{are} \\ \text{here} \\ \text{how} \\ \text{i} \\ \text{name} \\ \text{where} \\ \text{you} \end{bmatrix}$$

Take the hidden state vector to be $h_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$.

The trained weights and biases are given below:

$$W_h = \begin{bmatrix} -0.49 & -0.74 & -0.96 & -1.23 & -0.44 & 1.61 & -1.32 & 0.01 \\ -0.06 & 1.24 & -1.18 & 0.41 & -1.83 & -1.54 & -0.41 & -0.57 \end{bmatrix}$$

$$U_h = \begin{bmatrix} -3.78 & 6.39 \\ -6.25 & 1.71 \end{bmatrix}$$

$$b_h = \begin{bmatrix} -1.25 \\ 0.81 \end{bmatrix}$$

$$W_y = \begin{bmatrix} 2.91 & 6.41 \\ 0.08 & 1.16 \\ 5.36 & -3.93 \\ -0.18 & -0.4 \\ -5.55 & 1.38 \\ -0.49 & -0.64 \\ -0.81 & -0.14 \\ -2.21 & -0.18 \end{bmatrix}$$

$$b_y = \begin{bmatrix} -1.26 \\ -2.15 \\ 1.97 \\ -1.35 \\ 6.0 \\ -1.16 \\ -1.22 \\ -0.84 \end{bmatrix}$$

For the input sentence of "where are you", predict the output sentence using these weights and biases.

Solution

The input words are 'where', 'are' and 'you'. Based on the vocab vector, the corresponding one-hot encoded vectors (x_1, x_2 and x_3 respectively), are shown below.

$$x_1 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

$$x_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$x_3 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

For an RNN model, the equations are

$$h_t = \sigma\left(W_h x_t + U_h h_{t-1} + b_h\right)$$
$$y_t = s\left(W_y h_t + b_y\right)$$

For the first time step, we have

$$h_1 = \sigma\left(W_h x_1 + U_h h_0 + b_h\right)$$

$$\Rightarrow h_1 = \sigma\left(\begin{bmatrix} -0.49 & -0.74 & -0.96 & -1.23 & -0.44 & 1.61 & -1.32 & 0.01 \\ -0.06 & 1.24 & -1.18 & 0.41 & -1.83 & -1.54 & -0.41 & -0.57 \end{bmatrix}\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} -3.78 & 6.39 \\ -6.25 & 1.71 \end{bmatrix}\begin{bmatrix} 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -1.25 \\ 0.81 \end{bmatrix}\right)$$

$$\Rightarrow h_1 = \sigma\left(\begin{bmatrix} -1.32 \\ -0.41 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -1.25 \\ 0.81 \end{bmatrix}\right)$$

$$\Rightarrow h_1 = \sigma\left(\begin{bmatrix} -2.57 \\ 0.40 \end{bmatrix}\right) = \begin{bmatrix} 0.07 \\ 0.60 \end{bmatrix}$$

Hence, the output vector for the first time step would be

$$y_1 = s\left(W_y h_1 + b_y\right)$$

$$\Rightarrow y_1 = s\left(\begin{bmatrix} 2.91 & 6.41 \\ 0.08 & 1.16 \\ 5.36 & -3.93 \\ -0.18 & -0.4 \\ -5.55 & 1.38 \\ -0.49 & -0.64 \\ -0.81 & -0.14 \\ -2.21 & -0.18 \end{bmatrix}\begin{bmatrix} 0.07 \\ 0.60 \end{bmatrix} + \begin{bmatrix} -1.26 \\ -2.15 \\ 1.97 \\ -1.35 \\ 6.00 \\ -1.16 \\ -1.22 \\ -0.84 \end{bmatrix}\right)$$

$$\Rightarrow y_1 = s\left(\begin{bmatrix} 4.04 \\ 0.70 \\ -1.97 \\ -0.25 \\ 0.43 \\ -0.42 \\ -0.14 \\ -0.26 \end{bmatrix} + \begin{bmatrix} -1.26 \\ -2.15 \\ 1.97 \\ -1.35 \\ 6.00 \\ -1.16 \\ -1.22 \\ -0.84 \end{bmatrix}\right)$$

$$\Rightarrow y_1 = s\left(\begin{bmatrix} 2.78 \\ -1.45 \\ -0.00 \\ -1.60 \\ 6.43 \\ -1.58 \\ -1.36 \\ -1.10 \end{bmatrix}\right) = \begin{bmatrix} 0.03 \\ 0.00 \\ 0.00 \\ 0.00 \\ 0.97 \\ 0.00 \\ 0.00 \\ 0.00 \end{bmatrix}$$

If we convert this to a one-hot encoded vector with the maximum probability value converted to 1 and the rest of the values converted to zero, that would be $\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$, which corresponds to the word 'i'. Hence, the first output word is 'i'.

For the second time step, we have

$$h_2 = \sigma\left(W_h x_2 + U_h h_1 + b_h\right)$$

$$\Rightarrow h_2 = \sigma\left(\begin{bmatrix} -0.49 & -0.74 & -0.96 & -1.23 & -0.44 & 1.61 & -1.32 & 0.01 \\ -0.06 & 1.24 & -1.18 & 0.41 & -1.83 & -1.54 & -0.41 & -0.57 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -3.78 & 6.39 \\ -6.25 & 1.71 \end{bmatrix}\begin{bmatrix} 0.07 \\ 0.60 \end{bmatrix} + \begin{bmatrix} -1.25 \\ 0.81 \end{bmatrix}\right)$$

$$\Rightarrow h_2 = \sigma\left(\begin{bmatrix} -0.74 \\ 1.24 \end{bmatrix} + \begin{bmatrix} 3.56 \\ 0.58 \end{bmatrix} + \begin{bmatrix} -1.25 \\ 0.81 \end{bmatrix}\right)$$

$$\Rightarrow h_2 = \sigma\left(\begin{bmatrix} 1.57 \\ 2.63 \end{bmatrix}\right) = \begin{bmatrix} 0.83 \\ 0.93 \end{bmatrix}$$

Hence, the output vector for the second time step would be

$$y_2 = s\left(W_y h_2 + b_y\right)$$

$$\Rightarrow y_2 = s\left(\begin{bmatrix} 2.91 & 6.41 \\ 0.08 & 1.16 \\ 5.36 & -3.93 \\ -0.18 & -0.4 \\ -5.55 & 1.38 \\ -0.49 & -0.64 \\ -0.81 & -0.14 \\ -2.21 & -0.18 \end{bmatrix}\begin{bmatrix} 0.83 \\ 0.93 \end{bmatrix} + \begin{bmatrix} -1.26 \\ -2.15 \\ 1.97 \\ -1.35 \\ 6.00 \\ -1.16 \\ -1.22 \\ -0.84 \end{bmatrix}\right)$$

$$\Rightarrow y_2 = s\left(\begin{bmatrix} 8.39 \\ 1.15 \\ 0.77 \\ -0.52 \\ -3.30 \\ -1.00 \\ -0.80 \\ -2.00 \end{bmatrix} + \begin{bmatrix} -1.26 \\ -2.15 \\ 1.97 \\ -1.35 \\ 6.00 \\ -1.16 \\ -1.22 \\ -0.84 \end{bmatrix}\right)$$

$$\Rightarrow y_2 = s\left(\begin{bmatrix} 7.13 \\ -1.00 \\ 2.74 \\ -1.87 \\ 2.70 \\ -2.16 \\ -2.02 \\ -2.84 \end{bmatrix}\right) = \begin{bmatrix} 0.98 \\ 0.00 \\ 0.01 \\ 0.00 \\ 0.01 \\ 0.00 \\ 0.00 \\ 0.00 \end{bmatrix}$$

If we convert this to a one-hot encoded vector with the maximum probability value converted to 1 and the rest of the values converted to zero, that would be $\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$, which corresponds to the word 'am'. Hence, the second output word is 'am'.

For the third time step, we have

$$h_3 = \sigma\left(W_h x_3 + U_h h_2 + b_h\right)$$

$$\Rightarrow h_3 = \sigma\left(\begin{bmatrix} -0.49 & -0.74 & -0.96 & -1.23 & -0.44 & 1.61 & -1.32 & 0.01 \\ -0.06 & 1.24 & -1.18 & 0.41 & -1.83 & -1.54 & -0.41 & -0.57 \end{bmatrix}\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} + \begin{bmatrix} -3.78 & 6.39 \\ -6.25 & 1.71 \end{bmatrix}\begin{bmatrix} 0.83 \\ 0.93 \end{bmatrix} + \begin{bmatrix} -1.25 \\ 0.81 \end{bmatrix}\right)$$

$$\Rightarrow h_3 = \sigma\left(\begin{bmatrix} 0.01 \\ -0.57 \end{bmatrix} + \begin{bmatrix} 2.83 \\ -3.58 \end{bmatrix} + \begin{bmatrix} -1.25 \\ 0.81 \end{bmatrix}\right)$$

$$\Rightarrow h_3 = \sigma\left(\begin{bmatrix} 1.59 \\ -3.34 \end{bmatrix}\right) = \begin{bmatrix} 0.83 \\ 0.03 \end{bmatrix}$$

Hence, the output vector for the third time step would be

$$y_3 = s\left(W_y h_3 + b_y\right)$$

$$\Rightarrow y_3 = s\left(\begin{bmatrix} 2.91 & 6.41 \\ 0.08 & 1.16 \\ 5.36 & -3.93 \\ -0.18 & -0.4 \\ -5.55 & 1.38 \\ -0.49 & -0.64 \\ -0.81 & -0.14 \\ -2.21 & -0.18 \end{bmatrix}\begin{bmatrix} 0.83 \\ 0.03 \end{bmatrix} + \begin{bmatrix} -1.26 \\ -2.15 \\ 1.97 \\ -1.35 \\ 6.00 \\ -1.16 \\ -1.22 \\ -0.84 \end{bmatrix}\right)$$

$$\Rightarrow y_3 = s\left(\begin{bmatrix} 2.64 \\ 0.11 \\ 4.32 \\ -0.16 \\ -4.56 \\ -0.43 \\ -0.68 \\ -1.84 \end{bmatrix} + \begin{bmatrix} -1.26 \\ -2.15 \\ 1.97 \\ -1.35 \\ 6.00 \\ -1.16 \\ -1.22 \\ -0.84 \end{bmatrix}\right)$$

$$\Rightarrow y_3 = s\left(\begin{bmatrix} 1.38 \\ -2.04 \\ 6.29 \\ -1.51 \\ 1.44 \\ -1.59 \\ -1.90 \\ -2.68 \end{bmatrix}\right) = \begin{bmatrix} 0.01 \\ 0.00 \\ 0.98 \\ 0.00 \\ 0.01 \\ 0.00 \\ 0.00 \\ 0.00 \end{bmatrix}$$

If we convert this to a one-hot encoded vector with the maximum probability value converted to 1 and the rest of the values converted to zero, that would be $\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$, which corresponds to the word 'here'. Hence, the third output word is 'here'.

Hence, the predicted output sequence for the given input sequence of 'where are you' is 'i am here'.

Note: Computing softmax values is not really required here, because all that we need is the maximum value among the probabilities, which would be the same as the maximum value of the elements in the vector even before softmax is applied. This can save your time in exam.